Python warm up

Segmentation data

```
#meta=>genome_binning=1Mb;interval_number=3106
#group=>group_id=icdom-81403;label=Adenocarcinoma, NOS;dataset_id=progenetix;sam
group_id  chro  start end gain_frequency  loss_frequency  index
icdom-81403 1 0 1000000 8.8 9.12  0
icdom-81403 1 1000000 2000000 8.49  8.68  1
icdom-81403 1 2000000 3000000 9.81  13.19 2
icdom-81403 1 3000000 4000000 10.02 15.84 3
icdom-81403 1 4000000 5000000 7.94  15.91 4
...
icdom-81403 2 228000000 229000000 7.37  6.62  477
icdom-81403 2 229000000 230000000 7.39  6.89  478
icdom-81403 2 230000000 231000000 8.3 7.0 479
icdom-81403 2 231000000 232000000 8.24  6.86  480
icdom-81403 2 232000000 233000000 9.1 7.89  481
```

https://docs.progenetix.org/file-formats/#pgxfreq-segment-cnv-frequencies

Python warm up

- Data link: https://progenetix.org/beacon/variants/?output=pgxseg&filters=NCIT:C3030

- Check the data first, and write your own script to access and download the data via python.

- Transfer the data to dataframe in pycharm, with proper columns.

Python warm up

- **Histplot**: You can start by exploring the data to understand its structure and distribution. For example, you can check the distribution of the ‘reference_name' values using a histogram
- **Count plot**: Count the number of CNV events per biosample
- **Heatmap of CNV Events:** If you want to explore relationships between biosamples and CNV events, you can create a heatmap to visualize the presence or absence of CNV events across biosamples.

https://seaborn.pydata.org/generated/seaborn.histplot.html

https://doi.org/10.1093/database/baab043

- What is CNV/CNA?
- How will you describe or introduce progenetix (scale, data source, cancer types and so on)?
- Describe NCIt, ICOD, UBERON codes, and their relationships.
- What are CNV segmentations and CNV frequencies, and how to use them?
- What are APIs and how to use APIs in progenetix?
- How does progenetix visualise CNA profiles?
- What do you think should be improved in progenetix?

Please upload your file to your own folder of Bio392 GitHub, and name the file as lastname_firstname_paper_reading_day2.md.
It will be graded.

https://progenetix.org/

https://docs.github.com/en/get-started/writing-on-github/getting-started-with-writing-and-formatting-on-github/basic-writing-and-formatting-syntax